



**University of
Zurich^{UZH}**

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2012

Linked raster data

Scharrenbach, Thomas ; Bischof, Sandro ; Fleischli, Simon ; Weibel, Robert

Abstract: Linked Data (LD) is a paradigm that allows linking various heterogeneous data sources in a well-defined way (Bizer, Heath, Berners-Lee, 2009). These links establish a Web of Data (WoD), instead of a web of html documents. In the WoD more and more geographic data sets are made available such as GeoNames, dbPedia, etc. These are usually created from a GIS system and published as a static data set. Approaches like Geo-SPARQL aim towards processing these sets of Linked Geo Data but only little research has been performed how to use LD in the context of Geographic Information Systems (GIS)–both as an input to GIS operations as well as its output. Even less work has been done for LD-enabled raster data processing. Based on recent work (Fleischli, 2012) we investigate in this paper how geo-spatial data, in particular raster data and, even more important, operations on it, can be linked to the WoD in both directions.

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-74705>

Conference or Workshop Item

Published Version

Originally published at:

Scharrenbach, Thomas; Bischof, Sandro; Fleischli, Simon; Weibel, Robert (2012). Linked raster data. In: GIScience 2012: Seventh International Conference on Geographic Information Science, Columbus, Ohio, 18 September 2012 - 21 September 2012, online.

Linked Raster Data

Thomas Scharrenbach¹, Sandro Bischof², Simon Fleischli³, Robert Weibel³

¹University of Zurich, Department of Informatics, Binzmühlestrasse 14, CH-8050 Zurich, Switzerland
Email: scharrenbach@ifi.uzh.ch

²Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
Email: sandro.bischof@wsl.ch

³University of Zurich, Department of Geography, Winterthurerstr. 190, CH-8057 Zurich, Switzerland
Email: {simon.fleischli; robert.weibel@geo.uzh.ch

1 Introduction

Linked Data (LD) is a paradigm that allows linking various heterogeneous data sources in a well-defined way (Bizer, Heath, & Berners-Lee, 2009). These links establish a Web of Data (WoD), instead of a web of html documents. In the WoD more and more geographic data sets are made available such as GeoNames, dbPedia, OpenStreetMap, etc.¹ These are created from various sources such as, for example, Geographic Information Systems (GIS) and published as a static data set. Approaches like Geo-SPARQL² aim towards processing these sets of Linked Geo Data³ but only little research has been performed how to use LD in the context of GIS—both as an input to GIS operations as well as its output. Even less work has been done for LD-enabled raster data processing. Based on recent work (Fleischli, 2012) we investigate in this paper how geo-spatial data, in particular raster data and, even more important, operations on it, can be linked to the WoD.

2 Linked Data

Linked Data (Bizer et al., 2009) is a recommendation how to publish links to and between data entities, called resources. Following the RDF data model (Manola & Miller, 2004), links are established by *triples* $\langle s, p, o \rangle$ where p is a *qualified binary predicate* (aka *relation* or *property*) and o and s are *resources*. An IRI or an anonymous identifier (aka blank node) uniquely identifies each of these. As both the instance data as well as the data schema are expressed by RDF triples they can easily be linked by re-using common schemata or establishing links between different schemata. A data schema is preferably expressed in the W3C recommendations RDFS, OWL or SKOS, whereas data can be queried via SPARQL (Harris & Seaborne, 2012).

Using SPARQL, queries are represented as graph patterns that are matched against one or more RDF graphs. The results of a query are determined by evaluations of these graph patterns. Filters and join operations put constraints upon the evaluation and projections that determine the actual results. Results can be lists of n-tuples or RDF graphs. One of the distinct features of SPARQL is its ability to have a variable query predicate. The evaluation of the basic graph pattern `map1:poly1 ?prop map1:poly2`, for example, contains all links that exist between two particular polygons in a map.⁴ Evaluating the query variable `?prop`

¹ <http://geonames.org>; <http://dbpedia.org>; <http://linkedgeodata.org/>

² <http://geosparql.org/>

³ Note that the term Linked Geo Data is used in a general sense here. It is not to be mixed up with the LD version of OpenStreetMap.

⁴ We here assume that `map1:poly1` and `map1:poly2` refer to the prefixed IRI of two polygons.

will provide us with all those properties that exist between the two polygons, for example, topological relations expressed by `rcc:touches` or `rcc:tangentialProperPartOf`.

RDF triples can be universal or valid within a certain context only. In the first case, they are part of the *default graph* whereas they are part of a *named graph* in the latter case. Named graphs are referred to by an IRI and can be queried for, explicitly by using SPARQL. We will use named graphs to define the valid context of Linked Raster Data in Section 4.

3 Linked Geo-Data

Linked Data can help simplifying finding geographic entities. Consider, for example, the task of finding all those areas z_m of type hunting reserve (in German: “Jagdbanngebiet”) containing at least 20% forest and all intersecting municipalities x_n (cf Figure 1). LD enables meta-descriptions for geo-spatial entities such as layers and maps and potentially even single objects in a GIS. These meta-descriptions, in turn, are linked to other LD sources building a bridge from the geo-spatial domain to any domain of qualified typed information.

In our example, we would treat, for instance, each object z_m in the layer containing polygons of hunting reserves as an RDF resource and identify them with the type of a hunting reserve as defined by dbpedia: $z_m \text{ rdf:type dbpedia:Jagdbanngebiet}$.⁵ The description of a municipality in dbpedia links to other entities such as, for example, its German counterpart “Gemeinde” or the GML definition of the concept of a “Feature”. For these, in turn, there exist further links to other entities in other domains. The semantics of the polygons are now made explicit in the GIS, as a consequence.

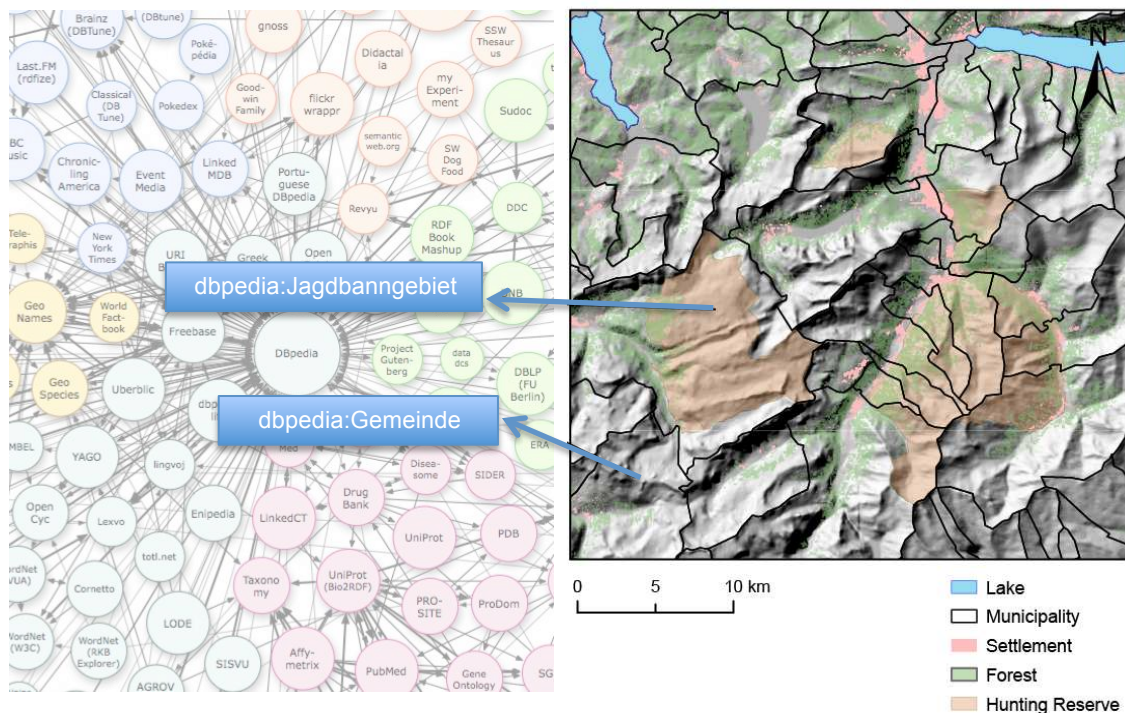


Figure 1: Linked Geo-Data for the intersection of several layers containing vector data (municipalities, lakes) and raster data (settlements, forests, hunting reserves). The arrows represent example RDF triples modeling the layers' semantics. On the left side part of the Linked Open Data cloud is presented to which the hunting reserves and the municipalities link (i.e. dbpedia).⁶

⁵ DBpedia is a LD representation of the entities in Wikipedia.

⁶ <http://richard.cyganiak.de/2007/10/lod/>

4 Linked Raster Data

Conceptually, raster data are used for representing *fields*: either *categorical fields* such as land cover categories, or *continuous fields* such as elevation, slope etc. While vector data are used to represent geographic entities and thus can be transformed directly to LD by assigning, for example, their objects an IRI, producing Linked Raster Data (LRD) is not as obvious.

There do not exist objects in a raster *per se* but only in a certain context. The ‘entities’ or ‘objects’ of a raster are sets of raster elements (i.e. sets of pixels) that meet a certain context-dependent definition. Examples would be actual real-world objects (*aka features*) such as rivers or roads, represented as connected components of pixels in a categorical raster; but also areas of pixels within the elevation band of 2100 to 2600 m.a.s.l. (the preferred altitude range for marmots in the Central Alps) or areas above 30° slope (the critical slope for snow avalanche formation). LRD can hence only refer to either *a) explicit features* (e.g. rasterized geo-objects) or *b) feature descriptions* (e.g. elevation band, slope class). As such we invariably project interpretations into the raster data. Needless to say that this is always the case when we form semantic categories in a raster/field, independently of the use of Linked Data.

Explicit features can be given a sound representation in RDF and GML, according to Usery & Varanka (2011). They use the example of Meteor Crater, Arizona to show how a bounded, areal geo-object can be represented in LD as sets of pixels of an orthoimage or scanned topo map. In contrast to their work, we also propose to define objects in terms of feature descriptions (FD). In Section 6 we will discuss the differences between the two cases.

Feature descriptions do not refer to actual objects but to *parameterized geographic operations*. This is of special interest for queries where the result depends on sub-queries or context information. The example of Section 3 already contains a feature description, i.e. the constraint that hunting reserves in the answer set must at least have 20% forest. Consider extending the example with the additional constraint that the resulting areas on average have to have a slope value that is larger than the average slope of all z_m . Clearly the answer set depends on a context and cannot be computed efficiently. Pre-computation would also not make sense because the constraint (above average slope) might change in a future query (e.g. to slope > 30°). Another important observation is that the answer set depends on the result of an operation on geo-objects.

Since objects resulting from feature descriptions are valid within a certain context only, *we store information about the objects in a unique named graph*. Conversely, information about the context is stored separately to be able to reference different contexts and, consequently, enable re-using contexts. Re-using context information or parts thereof simplifies caching of query results and allows defining new versions of those contexts, e.g. by using a different set of parameters.

5 GIS SPARQL Endpoint

In contrast to approaches like GeoSPARQL we do not propose having the geo-operations executed on the RDF representation. RDF refers to meta-information. As such, we only propose describing the semantics of geo-operations by defining their context and their parameters in RDF. Turning a GIS system with raster data into a SPARQL endpoint requires hence more than “just” assigning each object in the GIS an IRI and a type.

The actual process of matching graph patterns has to be adapted to incorporate the geo-operations implemented by the GIS. This can be achieved by combining the GIS’ algebra with that induced by SPARQL: We have to identify those GIS operations that directly operate on objects with RDF predicates. The evaluation of a basic graph pattern such as

`?x gis:contains ?y` would then contain the results of the operation in the GIS that `gis:contains` refers to.

6 Discussion

RDF focuses on modeling static meta-data. Yet, the nature of outputs from a GIS system is rather dynamic than static, as they result from (possibly parameterized) operations on the data.

In recent work we investigated several approaches towards Linked Raster Data Processing (Fleischli, 2012). It showed that for Linked Raster Data modeling *feature descriptions* has not been covered yet adequately for many GIS based use-cases. It was also indicated that modeling the single raster entries (i.e. the pixels) is not feasible but that a description of geo-operations leads to feasible and hopefully useful results.

Explicit features are better for cases where the feature extraction process cannot be described efficiently with RDF (e.g. for manually defined geo-objects or recursive object definitions). *Feature descriptions* are preferred otherwise. The actual choice depends on the given problem and the interpretations that we want to (and can) project into the data.

We did not address the representation of the actual data behind an IRI for raster data. This aspect is indeed well-discussed in (Usery & Varanka, 2011) and can be used for FD raster data, as well.

While we did not provide a detailed discussion, the methods discussed in this paper should be applicable to vector data without major effort. In the same manner we can provide LD descriptions for GIS operations on vector data. Yet, the results behind the links are objects (e.g. polygons) rather than raster data, which even simplifies the processing.

7 Conclusion and Outlook

In this paper we have advocated that LD has the potential to express the semantics of raster data used in a certain context. In addition to the existing approaches of representing actual explicit features (i.e. rasterized geo-objects such as lakes), we propose to describe the context of raster data in which it is valid (e.g. areas of slope above 30° as critical areas for avalanche formation). The context can consist of, for example, parameterized operations of a GIS and is encoded as Linked Data to facilitate caching and re-use.

Querying for objects represented as LRD feature descriptions we have to integrate GIS operations into the SPARQL graph pattern and filter evaluation. The results are finally represented as unique named graphs referring to their context in which they are valid.

The ability and shortcomings of representing the context of objects descriptions in LRD, and a detailed empirical evaluation of the usefulness of this approach remain for future work.

8 References

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems, 5(3), 1-22. Elsevier.
- Fleischli, S. (2012). Geodata as Linked Data -- Structuring and Linking of Environmental Data for the Semantic Web (unpublished). University of Zurich.
- Harris, S., & Seaborne, A. (2012). SPARQL 1.1 Query Language. W3C Technical Report, <http://www.w3.org/TR/sparql11-query/>
- Manola, F., & Miller, E. (2004). RDF Primer. W3C Technical Report, <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>
- Usery, E. L., & Varanka, D. (2011). Design and Development of Linked Data from The National Map. In: Semantic Web -- Interoperability, Usability, Applicability. Semantic Web Journal, IOS Press.